Crystallographic Structure of a DEAD-box RNA helicase

Jonathan M. Caruthers and David B. McKay

Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305

INTRODUCTION

Eukaryotic initiation factor 4A (eIF4A) is a member of a protein family referred to as DEA(D/H)-box RNA helicases. (The "DEA(D/H)" motif by which this family is named is the Walker "B" motif that participates in binding MgATP). These proteins participate in diverse activities that involve interactions with RNA; representative activities of these proteins include RNA splicing, ribosome biogenesis, and RNA degradation. Moreover, the activities of many proteins which are identified by sequence similarity as members of this family remain obscure. Our present subject, the 394-residue yeast eIF4A protein, is a prototype, minimal DEA(D/H)-box helicase. Previous work has provided the structure of the amino terminal, ATP binding domain of eIF4A [1, 2]. Extending those results, we have solved the structure of the carboxy terminal domain of eIF4A using multiwavelength anomalous scattering from a Sm^{3+} ion bound in the crystal. Using molecular replacement with the refined model of the carboxy terminal domain, we then completed the structure of the full-length eIF4A protein.

EXPERIMENTAL

The carboxy terminal domain was expressed and purified as described [3]. Crystals grew as elongated rods from 20% polyethylene glycol monomethyl ether of average molecular weight 550 (PEG-MME_{550}), 1mM $ZnSO_4$, 100mM MES pH 6.5; presence of Zn^{2+}, or a similar divalent metal ion, was mandatory for crystallization. For a heavy atom derivative, 2mM $SmCl_3$ was used instead of $ZnSO_4$ in the crystallization. For data collection, the concentration of PEG-MME_{550} in the mother liquor was increased to 40% for cryoprotection and crystals were flash-cooled in a stream of nitrogen gas at ~100 K. Native crystals are orthorhombic, space group $P2_12_12_1$, a=34.7 Å, b=52.1 Å, c=82.3 Å, with one molecule per asymmetric unit.

Native diffraction data were collected to a resolution limit of 1.75 Å on beamline 9-2 of the Stanford Synchrotron Radiation Laboratory (SSRL) at a wavelength of 0.980Å. Multiwavelength anomalous dispersion (MAD) data for a Sm^{3+} derivative were collected to a resolution limit of 3.0 Å at the Lawrence Berkeley Laboratory Advanced Light Source (ALS) synchrotron at wavelengths 1.84533 Å, 1.84616 Å, and 1.77120 Å for the Sm^{3+} absorption peak, dispersive edge and remote energies respectively. In both cases, data were recorded on Quantum IV CCD detectors and processed with the programs HKLVIEW, DENZO and SCALEPACK [4]. The location of a single Sm^{3+} ion was derived from anomalous difference Patterson peaks. MAD phases were computed using the program package CNS [5]; the overall figure of merit to 3.0 Å resolution was 0.80; phasing statistics are summarized in Table 1. After solvent flattening, much of the polypeptide chain could be traced in the experimental electron density map. Subsequent rounds of model refinement against the native data to 1.75 Å yielded the final model.

Table 1. Crystallographic data collection and multiwavelength phasing statistics. Sm^{3+} derivative of the carboxy terminal domain of yeast eIF4A. Resolution range (last shell), 26.6-3.0 (3.11-3.00) Å.

Wavelength	Observations (total/unique)	Completeness	R_{sym}*	<I>/<σ(I)>	Phasing power	f'	f''
λ1=1.7712Å	13729/5570	0.950 (0.945)	0.040 (0.056)	15.15 (15.02)	2.48	-12.1	10.9
λ2=1.8461Å	14174/5667	0.971 (0.980)	0.041 (0.062)	15.34 (13.93)	3.04	-36.1	15.2
λ3=1.8453Å	14176/5666	0.967 (0.954)	0.047 (0.069)	15.32 (14.03)	3.77	-21.5	35.8

	Resolution (Å)								
	5.96	4.74	4.14	3.77	3.5	3.29	3.13	2.99	Total
Reflections	641	700	672	701	708	741	716	659	5538
Figure of merit	0.810	0.833	0.775	0.798	0.794	0.776	0.795	0.764	0.799

Space Group: $P2_12_12_1$; unit cell (Å) a=34.7, b=52.1, c=82.3 (Å). Values of f' and f'' were initially estimated from an EXAFS scan and refined in CNS.
*$R_{sym} = \sum|I_{hkl} - \langle I_{hkl}\rangle|/\sum\langle I_{hkl}\rangle$ where I_{hkl} = single value of measured intensity of hkl reflection, and $\langle I\rangle$ = mean of all measured value intensity of hkl reflection.

RESULTS AND DISCUSSION

The carboxy-terminal domain has a parallel α−β structure (Fig. 1) with the same topology as the equivalent domain of distantly related DNA and viral RNA helicases. The conserved helicase sequence motifs IV, V and VI, as originally described by Gorbalenya and Koonin [6], are highlighted in Fig. 1. Their positions in the eIF4A domain are topologically equivalent to the positions of these motifs in other helicases In addition, we observe an arginine residue, Arg298 in helix 3, whose side chain is well-ordered in the structure, and whose spatial position suggests an involvement in RNA binding. Inspection of sequence alignments of DEA(H/D) RNA helicases (e.g. Pfam version 5.4, family 00271 [7]) reveal that this arginine is strictly conserved, and further, that it occurs as part of a QXXR motif, where the X amino acids are often charged. Notably, it is not present in the DNA helicases of SF-1, which lack this helix (and hence would not have been identified as a motif in sequence alignments looking for correlations across all helicases).

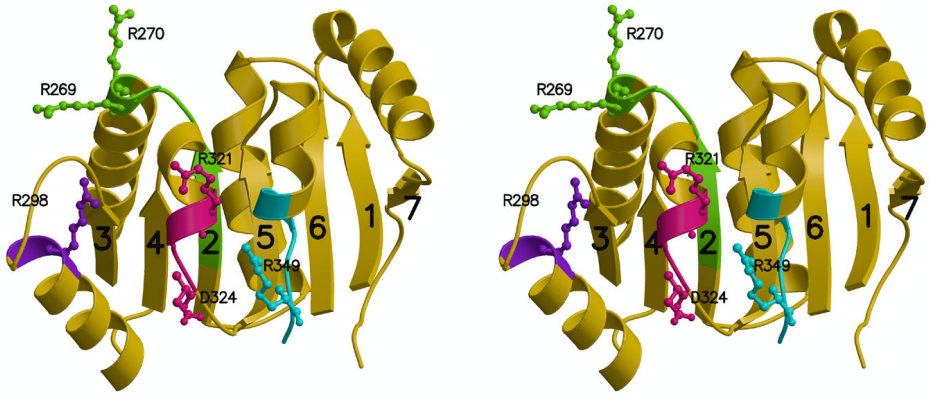

Figure 1. Stereo schematic drawing of carboxy terminal domain of yeast eIF4A.

Molecular replacement with the refined model of the carboxy terminal domain in our original crystal form of full-length eIF4A [1] shows that the molecule has a "dumbbell" structure in which the amino and carboxy terminal domains are connected by an extended 11-residue linker. The end-to-end length of the molecule is ~80 Å; the linker is ~18 Å long. The conserved helicase motifs on the amino and carboxy terminal domains have no spatial relationship that suggest a mechanism of coupling ATP binding/hydrolysis with RNA binding/release in this structure, suggesting that the molecule must undergo a significant conformational change in order to form a compact structure, in which the two domains interact with each other directly, at some point in the cycle which couples the enzymatic ATPase cycle to RNA duplex unwinding.

REFERENCES

1. E. R. Johnson and D. B. McKay, RNA **5**, 1526-1534 (1999).
2. J. Benz, H. Trachsel and U. Baumann, Structure **7**, 671-679 (1999).
3. J. M. Caruthers, E. R. Johnson and D. B. McKay, Proc. Natl. Acad. Sci. USA **97**, 13080-13085 (2000).
4. Z. Otwinowski and W. Minor, Meth. Enzymol. **276**, 307-326 (1997).
5. A. T. Brunger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson and G. L. Warren, Acta Cryst. D **54**, 905-921 (1998).
6. A. E. Gorbalenya and E. V. Koonin, Curr. Opin. Struct. Biol. **3**, 419-429 (1993).
7. A. Bateman, E. Birney, R. Durbin, S. R. Eddy, K. L. Howe and E. L. Sonnhammer, Nucleic Acids Res. **28**, 263-266 (2000).

This work was supported by a grant MCB-9874582 from the National Science Foundation to DBM, and Stanford Dean's Postdoctoral Fellowship to JMC.

Principal investigator: David McKay, Department of Structural Biology, Stanford University, Email: Dave.McKay@Stanford.edu, Telephone: 650-723-6589